Szegedi Orvostudományi Egyetem Számitástechnikai Központ
Szegedi Orvostudományi Egyetem Gyógyszertechnológiai Intézet

Gyógyszertartalmu porok keveredését leiró egyenlet kiválasztása statisztikai módszerekkel

Boda Krisztina, Eller József, Kata Mihály, Győri István

A szemcsés és porállományú keverékek megismerésének és tudományos alapon történő vizsgálatának szükségessége már a 40-es években felmerült. A gyógyszertartalmú porok technológiájának kutatása viszont csak mintegy 20 évvel ezelőtt indult meg. E keverési kisérletek konkrét gyakorlati feladatok megoldására irányultak és többnyire nem jutottak el az eredmények elméleti rendszerezéséig. Ma a porkeverés témájának különleges fontosságot ad az a tény, hogy a gyógyszertermékek 50-60 %-át szilárd gyógyszerformák képezik, amelyek készitésénél éppen az összetevők eloszlatása az egyik legfontosabb gyógyszertechnológiai müvelet.

A keverési kisérleteinkben 15 hatóanyagot és összesen 20 vivő és segédanyagot különböző arányokban tartalmazó keverékeket készitettünk. A keverésekhez Turbula készüléket használtunk 50, 70 és 90 percenkénti fordulatszámmal. Meghatározott időpontokban mértük a keverékek hatóanyagtartalmát.

Feladatunk a mérési adatok alapján a Turbula keverőben lejátszódó folyamat matematikai modellezése, a folyamatot leiró egyenlet meghatározása volt matematikai statisztikai módszerekkel.

A keveredés matematikai leirására szilárd heterogén rendszerek esetén /pl. gyógyszertartalmú porok/ általában az elsőrendü kémiai reakciók

$$y(t) = k(A - y(t))$$

egyenletét használták, melyben $y(t)$ jelenti a t időpontban átalakult anyagmennyiséget, A a kezdeti koncentráció /esetünkben = 100/, k a keveredési állandó. Az egyenlet megoldása $y(0) = 0$ kezdeti feltétellel

$$y(t) = A(1 - e^{-kt}) . \qquad (1)$$

Megkiséreltük ezt az egyenletet 20 előkisérlet mérési adatához illeszteni lineáris regresszió analizissel. A linearizálás az

$$\ln \frac{A}{A-y} = k.t$$

összefüggés alapján történt.

Ebben a 20 esetben a linearizálás után igen gyenge és nem szignifikáns korrelációkat kaptunk. Ez valószinüleg annak tudható be, hogy a fenti (1) egyenlet használata a folyamatnak a diffuzióval való hasonlóságán alapult [1].

Ugyanakkor a Turbula keverőben az anyagmozgás áramlásos jellegü.

Az illeszkedés jóságának növelése érdekében megkiséreltük ugyanezekhez a pontokhoz a másodrendü kémiai folyamatot leiró

$$y(t) = k(A - y(t))^2$$

differenciálegyenlet $y(0) = 0$ kezdeti feltételnek eleget tevő

$$y(t) = \frac{A^2kt}{1 + Akt} \qquad (2)$$

megoldását illeszteni. Mivel $A = 100$ rögzitett paraméter, a k-t kell becsülnünk.

Az illesztést most is transzformáció utáni lineáris regresszióval végeztük el. Kétféle transzformációt is kipróbáltunk a linearizálásra.

a./ Reciprok transzformáció

Ha feltesszük, hogy a mérési adatok jól illeszkednek a (2) egyenlethez, akkor a

$$t \rightarrow \frac{1}{t} \; ; \qquad y \rightarrow \frac{1}{y}$$

transzformáció végrehajtása után kapott

$$\frac{1}{y} = c \cdot \frac{1}{t} + b$$

lineáris összefüggés b rögzitett paraméterére $b = \frac{1}{A} \doteq \frac{1}{100}$ -nak kell teljesülnie. Ugyanis y-t kifejezve

$$y = \frac{\frac{t}{c}}{1 + \frac{b}{c} \cdot t}$$

egyenletet kapjuk.

b./ A másik transzformáció a $t \rightarrow t$; $y \rightarrow \frac{A.t}{y}$.
A transzformációt elvégezve

$$\frac{A \cdot t}{y} = c \cdot t + b$$

egyenletet kapjuk, átrendezve

$$y = \frac{\frac{A}{b} \cdot t}{1 + \frac{b}{c} \cdot t}$$

A rögzitett paraméter most c lesz, melyre c=1 -nek kell teljesülnie.

Mindkét transzformáció után a lineáris regresszió analizist elvégezve igen jó, szignifikáns korrelációkat kaptunk a transzformált adatokra vonatkozóan, és a rögzített paraméter értékek is elég pontosan adódtak.

Ezek alapján úgy láttuk, hogy a folyamat leirására a másodrendü egyenlet alkalmas.

Előkisérleteink tehát a másodrendü modell mellett szóltak, bár a k paraméter lineáris transzformáció utáni becslését nem éreztük pontosnak.

Ezért további 150 kisérletből nyert méréseken már nemcsak a transzformáció utáni lineáris regresszió-analizist végeztük el.

Ezekre a görbékre nemlineáris legkisebb négyzetes paraméterbecslő eljárást ill. egy Box és Cox által leirt módszer modell-diszkriminációra használt változatát is alkalmaztuk.

A nemlineáris legkisebb négyzetes módszer, amellyel a k paramétert becsültük, a Newton-féle iteráción alapult. Az iterációhoz jó kezdeti értékül szolgált a linearizált modellre kapott becslés.

Az ilyen módon meghatározott $\hat{k}$ értékek alapján kiszámitottuk a legjobban illeszkedő

$$\hat{y}_i = \frac{A^2\hat{k}t_i}{1 + A\hat{k}t_i}$$

értékeket. Az illeszkedés mértékéül az

$$I_m = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i y_i^2}}$$

relativ négyzetes hibát választottuk.

A modell-diszkriminációra alkalmazott módszer alapja Box és Cox által 1964-ben közölt módszer [2], amely feltételes maximum likelihood becslésen alapul. Ezt a módszert most nem kivánjuk részletesen ismertetni, mivel a szerzőkön kivül még pld. Kendall [3] is részletesen tárgyalja.

Box és Cox módszerét modell-diszkriminációra Sclove [4] használta először 1972-ben. Az ő módszerét a mi esetünkben mutatjuk be.

A "λ-rendü" folyamat egyenlete:

$$\dot{y}(t) = k(A - y(t))^{\lambda}$$

$y(0) = 0$ kezdeti feltétellel vett megoldása:

$$y(t) = A - (A^{1-\lambda} + (\lambda-1)kt)^{\frac{1}{1-\lambda}}, \qquad (\lambda \neq 1)$$

melynek linearizált alakja

$$y_{\lambda} = k.t$$

ahol

$$y_\lambda = \begin{cases} \dfrac{(A-y)^{1-\lambda} - A^{1-\lambda}}{\lambda - 1} & \lambda \neq 1 \\ \log \dfrac{A}{A-y} & \lambda = 1 \end{cases}$$

Feltételezzük, hogy az (y^i, t^i) mérési pontokra $\lambda=1$ vagy $\lambda=2$ esetén a következő összefüggés teljesül:

$$y_\lambda^i = k.t^i + \varepsilon^i \qquad i=1,\ldots,n$$

ahol ε^i normális eloszlású 0 várható értékkel, δ_λ^2 szórásnégyzettel és ε^i bármely ε^j-vel korrelálatlan $i \neq j$ esetén. Az y^i sürüségfüggvénye:

$$f_i(\lambda) = \frac{1}{\sqrt{2\pi\delta_\lambda^2}} \, e^{-\frac{(y_\lambda^i - k.t^i)^2}{2\delta_\lambda^2}} \, (A-y^i)^{-\lambda}$$

amit felhasználva az eredeti megfigyelések likelihood függvénye:

$$L(\lambda) = \prod_{i=1}^{n} f_i(\lambda)$$

Ekkor

$$\ell(\lambda) = \log L(\lambda) = -\frac{n}{2}\log(2\pi\delta_\lambda^2) - \sum_{i=1}^{n} (y_\lambda^i - k.t^i)^2 / 2\delta_\lambda^2$$

$$- \lambda \sum_{i=1}^{n} \log(A-y^i)$$

Fix λ-ra a $k = k(\lambda)$ és a δ^2_λ maximum likelihood becslése

$$\hat{k} = \hat{k}(\lambda) = \frac{\sum_{i=1}^{n} y^i_\lambda t^i}{\sum_{i=1}^{n} t_i^2}$$

$$\delta^2_\lambda = \frac{S_\lambda}{n} \; ; \qquad S_\lambda = \sum_{i=1}^{n} (y^i_\lambda - \hat{k}t^i)^2$$

Ezeket a becsléseket figyelembe véve fix λ-ra a log likelihood függvény maximális értéke:

$$\ell_{max}(\lambda) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log S_\lambda + \frac{n}{2}\log n - \frac{n}{2} - \lambda\sum_{i=1}^{n}\log(A-y^i)$$

Minket elsősorban az érdekel, hogy milyen λ-ra lesz ez a kifejezés maximális $(\lambda \in \{1,2\})$.

Átrendezve aszerint, hogy mely tagok tartalmazzák λ-t és melyek nem, a következő kifejezést kapjuk, melyet $T(\lambda)$-val jelölünk:

$$T(\lambda) = -\left[\ell_{max}(\lambda) - \frac{n}{2}\log n + \frac{n}{2}\right] = \frac{n}{2}\log S_\lambda + \lambda\cdot\sum_{i=1}^{n}\log(A-y^i)$$

$T(\lambda)$ az $\ell_{max}(\lambda)$ -nak csökkenő függvénye: ha $T(\lambda) < T(\lambda')$, akkor a λ-nak megfelelő transzformáció jobbnak látszik, mint a λ'-nak megfelelő.

Esetünkben

$$T(1) = \frac{n}{2} \log S_1 + \sum_{i=1}^{n} \log (A-y^i) ,$$

$$T(2) = \frac{n}{2} \log S_2 + 2 \sum_{i=1}^{n} \log (A-y^i) .$$

Azt a transzformációt fogadjuk el jobbnak, amelyre a $T(\lambda)$ értéke kisebb.

Ennek a módszernek az esetünkben történő alkalmazását arra alapoztuk, hogy Sclove a módszerét nagyszámú - bár más transzformációval nyert adaton - kipróbálta. A szimulációs vizsgálataival alátámasztotta a módszer robosztus voltát, azaz, hogy a módszer kevéssé érzékeny a hiba eloszlás normálistól való eltérésére. Összehasonlitó vizsgálatokat tett továbbá arra vonatkozóan, hogy a korrelációs együttható alkalmazása vagy az általa definiált $T(\lambda)$ mérőszám használata-e a jobb a két modell közötti választás szempontjából. Az esetek többségében nem talált nagy különbséget, de bizonyos paramétertartományokban a korrelációs együtthatón alapuló döntés rosszabb volt a likelihood módszerre alapozott döntéshez képest.

Eredményeink

150 keverés adatain lefuttattuk a linearizált regressziós modellt, a legkisebb négyzetes közelitést a linearizált modellből kapott kezdeti értékekkel és a Box és Cox módszerét.

MÓDSZER	Elsőrendü kémiai folyamat egyenlete	Másodrendü kémiai folyamat egyenlete
Nemlineáris paraméter-becslés $I_m=\sqrt{\dfrac{\sum_i (y^i - \hat{y}^i)^2}{\sum_i {y^i}^2}}$	0 $I_m(1) < I_m(2)$	149 $I_m(1) > I_m(2)$
Maximum likelihood	4 $T(1) < T(2)$	145 $T(1) > T(2)$

A legkisebb négyzetes közelitésnél használt relativ négyzetes hiba mindig kisebb volt másodrendü esetben.

A $T(\lambda)$ érték is az esetek nagy részében a másodrendü egyenlet esetén volt jobb /azaz kisebb/: 149-ből 145 esetben. Abban a 4 esetben viszont, mikor a $T(\lambda)$ az elsőrendü esetben volt kisebb, a korrelációs együttható viszont a linearizált modellnél a másodrendüre döntött.

Mindez alapján azt állithatjuk, hogy a Turbula keverőben gyógyszertartalmú porok keveredésének matematikai leirására az eddig használt elsőrendü kémiai reakcióegyenlet helyett a másodrendü látszik alkalmasabbnak. Ehhez az állitáshoz statisztikai módszerek - legkisebb négyzetes paraméterbecslés, likelihood módszer - segitségével jutottunk el.

A k paraméterre kapott paraméterbecslést gyógyszertechnológiai szempontból jól fel lehetett használni a keverési állandó optimális értékének megállapitására, valamint annak jellemzésére, hogy a keverés eredményes volt, átmeneti vagy rossz eredményü. Állitásunk a gyógyszertechnológia elmélete szempontjából új megállapitást jelent.

I r o d a l o m

[1] Train, D.: J. Am. Pharm. ASS. Sci. Ed. 49. 265. /1960/.

[2] Box, G. E. P., Cox, D. R. /1964/: "An Analysis of Transformations" I. R. Statist. Soc. Series B, 26, 211-243.

[3] Kendall, M. G., Stuart: The Advanced Theory of Statistics Vol. 2. Griffin London, 1973.

[4] S. L. Sclove: (Y vs. X) or (log Y vs X)? Technometrics Vol. 14. No. 2., May, 1972. 391-403.